# The visual perception of relative mass from object collisions

**James T. Todd** Ohio State University, Columbus, OH, USA

**J. Farley Norman** Western Kentucky University, Bowling Green, KY, USA

**Previous research has shown that observers can make reliable judgments about the relative mass of moving objects that collide in animated displays. One popular explanation of this is that observers' judgments are based on an internal model of Newtonian dynamics. An alternative explanation is that these judgments are based on measurable optical properties that are correlated with relative mass. To better understand this issue, the present investigation reanalyzed the data from three previous studies by Mitko and Fischer (2023), Sanborn et al. (2013), and Todd and Warren (1982), and it replicated an additional study by Hamrick et al. (2016). These new analyses demonstrate that observers' judgments of relative mass are most likely based on the post-collision optical velocities of objects without having to invoke an implausible mental representation of Newtonian dynamics as has been argued by several previous investigators.**

## Introduction

Human beings must continually deal with the laws of physics to survive in the natural environment. The most obvious examples of this involve motor control processes such as balancing on two legs or softly grasping a delicate object. However, the laws of physics can also have an important influence on visual perception. The visible behaviors of interacting objects are governed in part by their intrinsic mechanical properties like mass, rigidity, elasticity, or viscosity, and the effects of those properties are revealed perceptually by systematic variations within patterns of optical motion.

There are two competing theoretical models about how it is possible for human observers to make perceptual judgments about the mechanical properties of objects. One approach assumes that observers have an internal model of physics that can simulate the behavior of objects and events in the natural environment using equations derived from classical Newtonian dynamics. An alternative sensory based approach attempts to explain judgements of mechanical properties based on optical information that covaries with those properties in the natural environment. In general, this optical information will be only partially correlated with mechanical properties, which can produce systematic errors in observers' judgments. In the present article we will examine several previous experiments on the perception of mechanical properties in an effort to determine which approach provides a better account of human perception. We will also report a new replication of one of those experiments to more carefully explore the available optical information.

## Relative mass: A test case of perceived mechanical properties

Consider the images of two colliding objects that are depicted in Figure 1. This type of event was first studied by Michotte (1946/1963) in his pioneering research on the perception of causality. It was later noted by Runeson (1977) that there is potential visual information in these displays to specify the relative masses of the two objects based on the law of conservation of momentum. Suppose that there are two objects, a and b, that undergo a head-on collision, where $U_a$ and $U_b$ are their respective velocities before the collision; $V_a$ and $V_b$ are their respective velocities after the collision; and $M_a$ and $M_b$ are their respective masses. From the law of conservation of momentum, it follows that:

$$M_a/M_b = (U_b - V_b) / (V_a - U_a). \quad (1)$$

The elasticity of the collision is measured by the coefficient of restitution (e) as defined by:

$$e = |V_b - V_a| / |U_a - U_b|. \quad (2)$$

In other words, the elasticity of the collision and the mass ratio of the objects are optically specified by their relative velocities before and after the collision.

The first psychophysical investigation of Runeson's analysis was performed by Todd and Warren (1982).

Citation: Todd, J. T., & Norman, J. F. (2026). The visual perception of relative mass from object collisions. *Journal of Vision*, *26*(2):1, 1–15, https://doi.org/10.1167/jov.26.2.1.

https://doi.org/10.1167/jov.26.2.1 Received July 7, 2025; published February 2, 2026 ISSN 1534-7362 Copyright 2026 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from jov.arvojournals.org on 02/02/2026

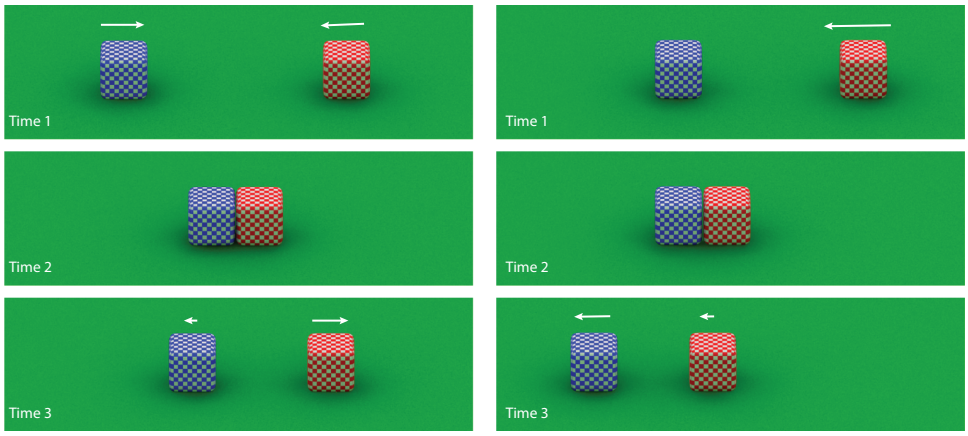


Figure 1. Two types of collisions from Todd and Warren (1982). In both cases, the blue object is 50% heavier than the red object and e = 0.5. When the objects approached each other with the same speed (left), the observers were 95% accurate. However, when a moving object collided with a stationary one (right), the observers were only 23% accurate. Videos of these events can be observed in Animation 1 of the supplementary materials.

Observers viewed computer animations of colliding objects (e.g., see Figure 1), and they were asked to report which object appeared heavier. When both objects moved at the same speed in opposite directions (the moving condition), observers' judgments were almost 100% accurate. However, when one object was initially stationary (the stationary condition), there was a strong bias to perceive the moving object as heavier. This effect has been replicated by numerous other investigators (e.g., Cohen, 2006; Deeb & Domini, 2025; Gilden & Proffitt, 1989; Mitko & Fischer, 2023; Sanborn, Mansinghka, & Griffiths, 2013), and it is now referred to as the *moving object bias*.

Another strong prediction of Runeson's model is that judgments of relative mass should be invariant over variations in elasticity, but that prediction was not confirmed by the data. The overall performance deteriorated significantly as the elasticity was reduced, and that result has also been replicated by Sanborn et al. (2013). Based on these findings, Todd and Warren (1982) speculated that observers' judgments may not involve an explicit computation of relative mass as described by Equation 1 but may instead be determined by optical variables that are only partially correlated with relative mass (see also Gilden & Proffitt, 1989).

Sanborn et al. (2013) have proposed a Bayesian analysis for estimating the relative masses of colliding objects that they refer to as a "Noisy Newton" model. It uses the same equation proposed by Runeson (1977), but it allows for the presence of noise, and prior expectations for the distributions of object velocities, mass, and the coefficient of restitution. Sanborn et al. (2013) also replicated the earlier study of Todd and Warren (1982). The filled symbols in the left panel of Figure 2 show the percentage of trials in the stationary condition where the moving object was judged as heavier for different values of relative mass and the coefficient of restitution. Note that there is a strong bias to perceive the moving object as heavier and a significant effect of elasticity. The solid lines in that figure show the predictions of the noisy Newton model. It is important to keep in mind that the underlying equation of this model cannot account for either of these effects. That is achieved by adding numerous free parameters (i.e., priors) to modify the output of that equation. These are implemented as prior distributions for mass, velocity, and the coefficient of restitution. Although this model predicts a moving object bias and is highly correlated with the data ($r = 0.93$), it systematically underestimates the percentage of moving object responses across all conditions by an average of 9.5%.

Are relative mass judgments really based on the conservation of momentum as described by Equation 1? Gilden and Proffitt (1989) performed an interesting experiment that relates directly to this issue. They demonstrated that observers are able to judge the relative mass of colliding objects even when all motion prior to the moment of contact is masked from view. That should be impossible according to the models of Runeson (1977) or Sanborn et al. (2013), and it provides strong evidence that observers relied on some other source of information based solely on the post-collision velocities.

The right panel of Figure 2 shows how the data from Sanborn et al. (2013) are correlated with the sum of the post-collision velocities in each display. A four-parameter logistic regression analysis was performed to compute the best fitting sigmoidal function for the entire data set. The solid lines in this figure show the predictions of that function for each of the three elasticity conditions. Note

Downloaded from jov.arvojournals.org on 02/02/2026

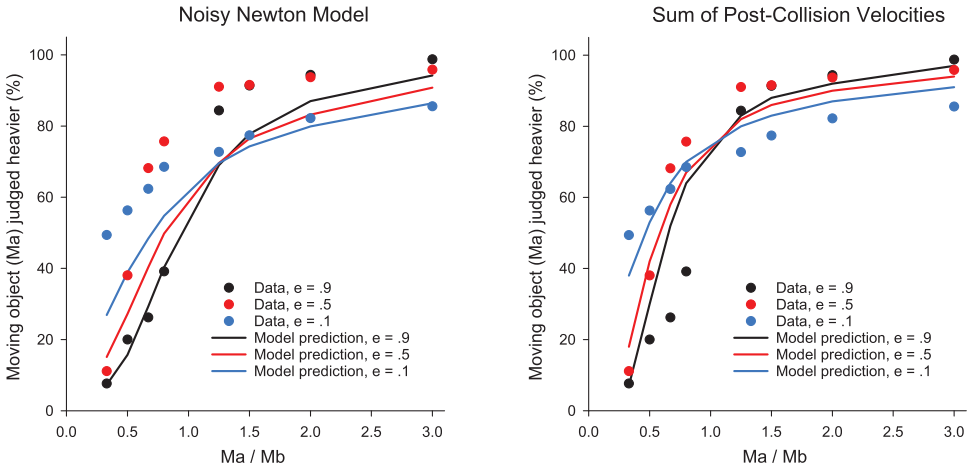


Figure 2. Left: The filled circles show relative mass judgments obtained by Sanborn et al (2013) when a moving object collided with a stationary one under multiple conditions with variable mass ratios and coefficients of restitution (e). The solid curves show the predictions of the noisy Newton model for each possible value of e. Right: The filled circles are the same as in the left panel, but the curves show the predictions of an alternative model based on the sum of the post-collision velocities.

that this model produces a marginally better fit to the data than the noisy Newton model ($r = 0.95$) with a smaller number of free parameters, and the average error of the predicted responses was only 1.4%. Similarly good fits were also obtained using this method for the data of Todd and Warren (1982).

Of course, there are other factors for evaluating models in addition to goodness of fit. The noisy Newton model might be considered more plausible because it is based on actual physics, whereas the alternative model seems relatively ad hoc. However, a more careful examination reveals that the specific parameters employed by the noisy Newton model have little or no physical justification. For example, the model assumes that the masses of objects follow an exponential distribution, such that lower masses are more likely than higher ones. This would imply that objects with zero mass are observed most frequently in the natural environment, which is obviously not the case. Similarly, the model assumes that the coefficient of restitution has a uniform prior distribution over its entire range from 0 to 1, which also has no ecological justification. For example, a value of one is physically impossible because it would allow a bouncing ball to remain in perpetual motion. Although not stated explicitly, the noisy Newton model shares two other assumptions with Runeson's analysis that are seldom, if ever, satisfied for real world collisions: They both assume that objects slide along a frictionless surface without any spin.

A closer examination also reveals that the sensory based model shown in the right panel of Figure 2 is not as ad hoc as may appear at first blush. When a moving object collides with a stationary one, the amount of energy in the system varies proportionally with the mass of the moving object. As the mass of that object is increased relative to the stationary one, both objects will move faster after the collision. Similarly, as the elasticity of the collision is increased, more of the total energy is preserved, and both objects will move faster after the collision. Note that this is a general constraint on post collision velocities that does not depend on other factors like friction or spin. There are two sources of evidence to indicate that observers' judgments of relative mass may be based on this constraint: One is the finding of Gilden and Proffitt (1989) that observers can judge the relative mass of colliding objects even when all motion prior to the moment of contact is masked from view. The other is the high correlation of observers' judgments with the post collision velocities in the experiments of Todd and Warren (1982) and Sanborn et al. (2013), as shown in the right panel of Figure 2.

## The effects of friction

When one object slides over another in the natural environment there is a frictional force in the opposite direction that causes the sliding object to decelerate. The magnitude of this force is determined by the mass of the sliding object and the coefficient of kinetic friction. For example, when a glass object slides over a glass surface, the coefficient of kinetic friction is 0.4. If the two surfaces are lubricated, the coefficient can be reduced considerably to around 0.1, but it can never be eliminated entirely.

Some simulations of low friction collisions between a moving red object and a stationary blue one are shown

Downloaded from jov.arvojournals.org on 02/02/2026

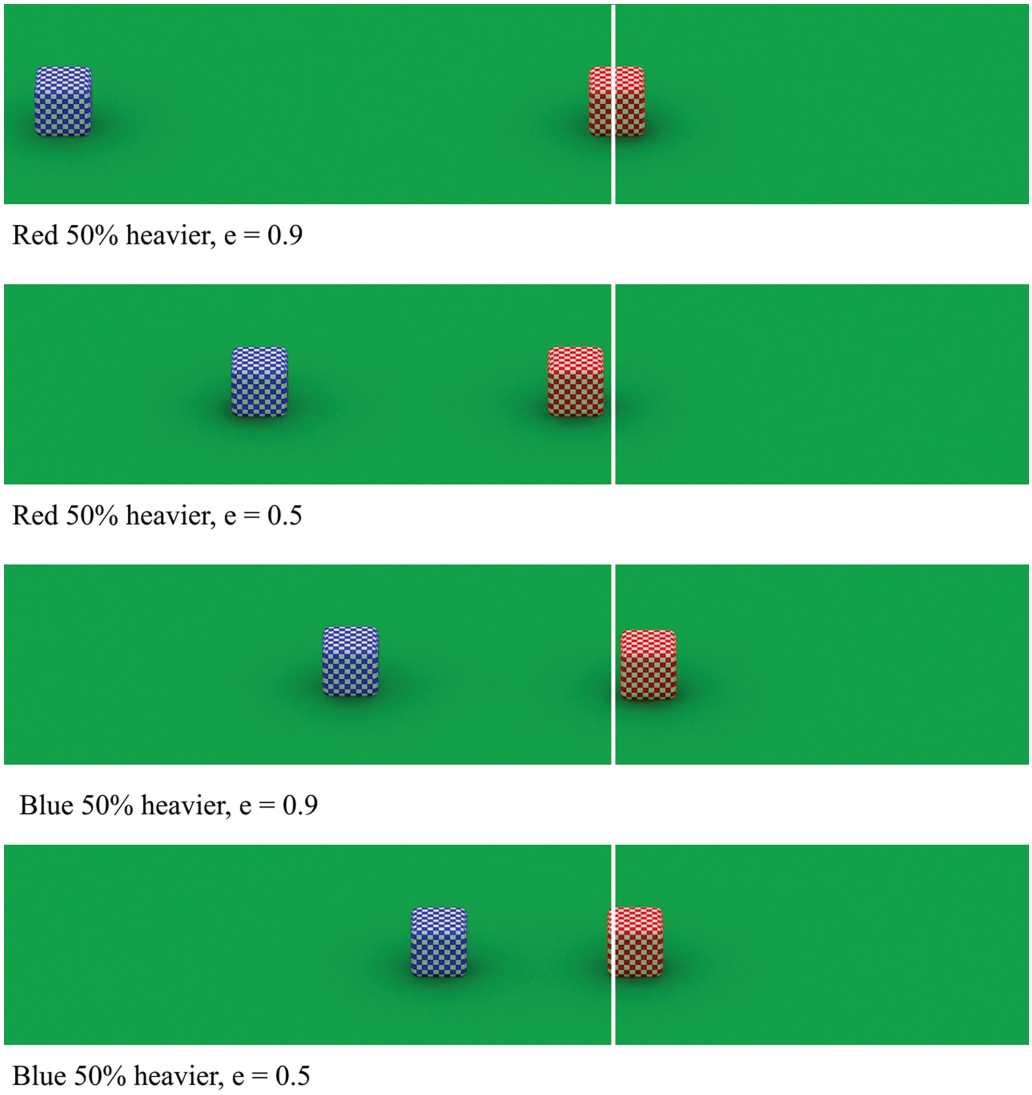


Figure 3. The final resting positions when a moving red object collides with a stationary blue object with varying values of e and relative mass. The simulated coefficient of friction is 0.1, and the initial velocity of the red object is the same in all conditions. The point of contact between the two objects is indicated by a white line. Videos of these events can be observed in Animation 2 of the supplementary materials.

in Figure 3, and videos of those events can be observed in Animation 2 of the supplemental materials. The vertically stacked panels show the final resting positions of both objects for different values of e and relative mass. The initial velocity of the red object is the same in all conditions, and the point of contact between the two objects is indicated by a vertical white line. The main thing to note in the animations is that the objects always decelerate after the collision until they come to a complete stop.

Note that the equations used by Runeson (1977) and Sanborn et al. (2013) are not applicable under these conditions, because both objects decelerate rapidly after the collision. However, the general principle about post-collision velocities described above is still valid. The amount of energy in the system varies proportionally with the mass of the moving object, but, in this case, almost all of that energy is transferred to the initially stationary object. As the mass of the moving object is increased, the stationary one will move faster and farther after the collision. Similarly, as the elasticity of the collision is increased, more of the total energy is preserved, and the stationary one will also move faster and farther. Thus, the post-collision movement of the stationary object provides useful information about relative mass, although that information is confounded with the effects of elasticity, as is also the case for frictionless surfaces.

Downloaded from jov.arvojournals.org on 02/02/2026

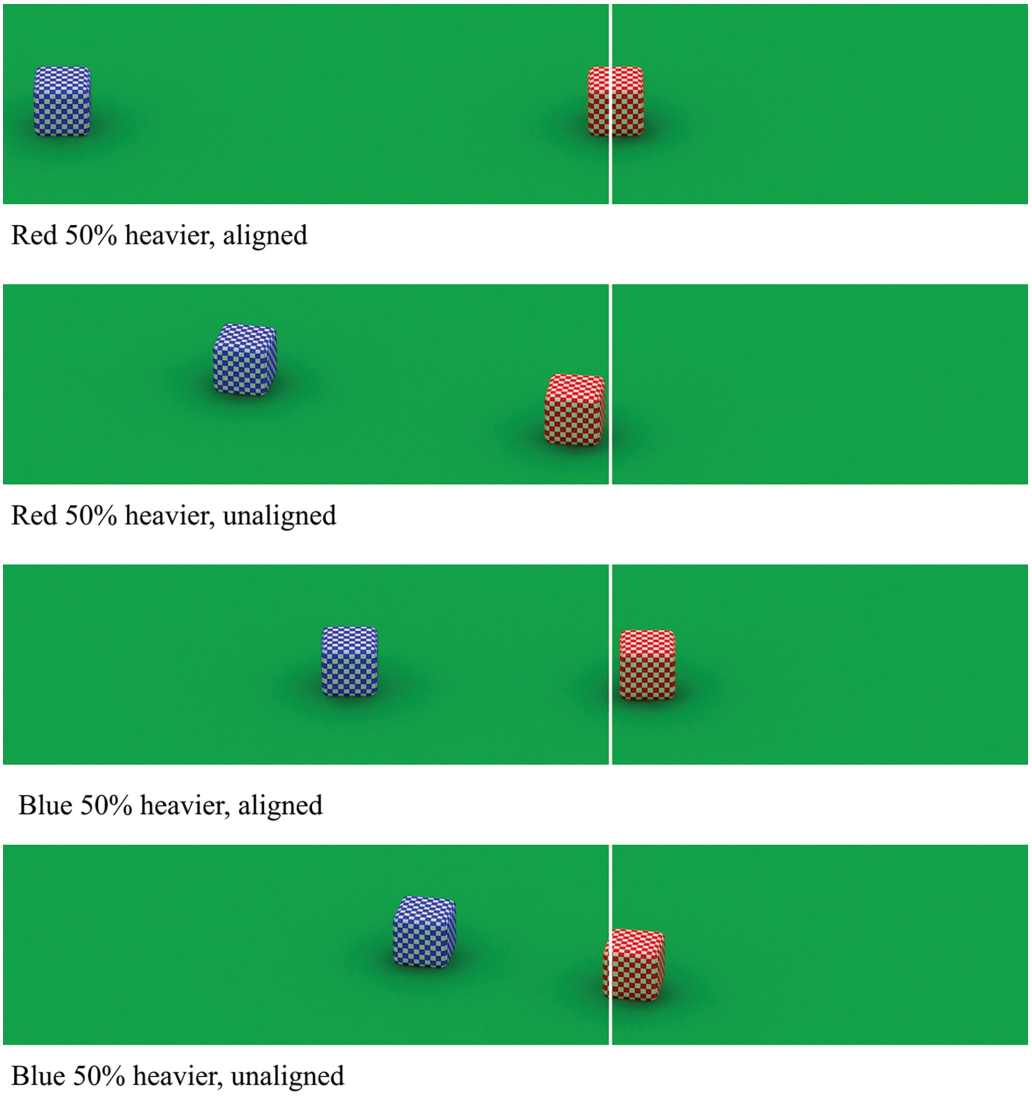


Figure 4. The final resting positions when a moving red object collides with a stationary blue object. The objects can be aligned or unaligned at the point of contact, which is indicated by a white line. The initial velocity of the red object is the same in all conditions. Videos of these events can be observed in Animation 3 of the supplementary materials.

## The effects of spin on sliding blocks

All of the examples of colliding objects presented thus far depict cubes sliding over a planar surface, and those cubes have been perfectly aligned at the point of contact. This alignment has important consequences for what happens next. When two objects collide, they each exert a force on one another. If two colliding cubes are perfectly aligned, then that force is distributed evenly over the two faces that make contact with one another. However, if the objects are misaligned, the forces will be unevenly distributed. This will create a torque that causes both objects to spin about their respective centers. It is important to keep in mind that the total amount of kinetic energy can never increase as the result of a collision. Thus, if some of that energy is converted to spin, then the linear displacements of the objects must necessarily be reduced.

Figure 4 shows examples of some aligned and unaligned collisions between a moving red object and a stationary blue one with different mass ratios. The vertically stacked panels show the final resting positions after both objects have come to a complete stop. The initial velocity of the red object is the same in all conditions, and the point of contact between the two objects is indicated by a white line. It is important to emphasize that the probability of a perfectly aligned collision in the natural environment is vanishingly small. Thus any perceptual computation of relative mass based on Newton's laws would need to consider the

Downloaded from jov.arvojournals.org on 02/02/2026

possible effects of spin. However, the general rule about post-collision velocities is still valid. The post-collision movement of the stationary object provides useful information about relative mass, which is confounded in this case by the effects of spin.

## The effects of spin on rolling balls

It is interesting to note that some previous studies of the perception of collision events have used objects such as cubes with a flat bottom (e.g., Todd & Warren, 1982; Sanborn et al., 2013), whereas others have employed spherical objects (e.g., Gilden & Proffitt, 1989; Mitko & Fischer, 2023). This makes little difference in a frictionless environment, but that is not the case under more natural circumstances. The effects of friction are much less on spherical objects because they only contact the ground at a single point. Moreover, the friction that does occur creates a torque that causes the object to roll forward. On a frictionless surface a ball would slide rather than roll.

It is somewhat surprising that the effects of spin have been largely ignored in the literature on perceived dynamic events, given the importance of these effects in a wide variety of games that involve moving balls, such as bowling, golf, tennis, or baseball. For example, consider the game of billiards, in which a player strikes a ball with the tip of a stick (i.e., a cue) to make it collide with another ball, so that the second ball will roll into a pocket. A skilled player can manipulate the spin on the first ball by striking it at different locations relative to its center to set up the next shot. Striking the ball above the center creates topspin that causes the ball to roll forward after the collision. Striking below the center creates backspin that causes the ball to roll backward after the collision. The ball can also be manipulated to roll left or right by striking it at a lateral position relative to its center.

Mitko and Fischer (2023) have recently studied the perception of relative mass from the collisions of real bowling balls, which varied in weight between six and 16 pounds. One ball was set in motion by rolling down a ramp onto a flat surface where it collided with a stationary ball. Observers viewed videos of these events and were asked to judge which ball appeared heavier. Figure 5 shows still images at different points in time before and after a collision for balls with varying mass ratios, and videos of these events can be observed in Animation 4 of the supplementary materials. These animations were graciously provided to us by Jason Fischer. Although not visible in the still images, the animations of these events reveal a complex interaction between friction and spin. Note in the upper panels where the moving ball is lighter that it first slides backward after the collision, although it spins in the opposite direction. Once the spin gains traction, however, the ball reverses its direction and rolls forward.

Note also that the general principle about post-collision velocity is clearly evident in this figure. That is to say, the depicted displacement of both objects varies systematically with the relative mass of the two objects. As the mass ratio between the moving and stationary object increases, the displacements of both objects become systematically larger, which could provide useful information for human perception. To test that possibility, we estimated the average post-collision velocity of the initially stationary object for all of the Mitko and Fischer stimuli by measuring the number of video frames after the collision until the object disappeared from view. These estimated velocities were almost perfectly correlated with the mass ratio of the two depicted bowling balls ($r = 0.96$). We then correlated those velocities using logistic regression with the proportion of trials that the moving object was judged to be heavier for each stimulus in their experiment. The results of this analysis are shown in Figure 6. The average post-collision velocity of the stationary object accounts for 96% of the variance in their experiment without having to make any assumptions about prior distributions of physical variables, like mass, elasticity, spin, or friction.

It is also important to note in this figure that there was a clear moving object bias in the average data that is similar to the results of previous investigations using simulated events without friction or spin (e.g., Cohen, 2006; Deeb & Domini, 2025; Gilden & Proffitt, 1989; Sanborn et al., 2013; Todd & Warren, 1982). The trials where the moving object was heavier (black dots) were judged more accurately than the trials where the stationary object was heavier (red dots). However, Mitko and Fischer found that there were huge individual differences in the point of subjective equality (PSE), which spanned a range of mass ratios from 0.4 to 1.5 (See Figure 3a of Mitko & Fischer, 2023). Note that the PSE values greater than 1.0 indicate a bias to judge the stationary object as heavier, and 27% of the observers fell into that category.

If observers' judgements were based on the post-collision velocities of the objects, they would need to adopt some setpoint to determine a PSE for the two colliding objects. If different observers adopted different set points, it would cause their psychometric functions to be shifted relative to one another, which could account for these large individual differences. There is also some empirical evidence to suggest that the PSEs may be modified if observers are given feedback after each trial (Jacobs, Michaels, & Runeson, 2000; Jacobs, Runeson, & Michaels, 2001; Runeson & Andersson, 2007).

It is unlikely that the noisy Newton model of Sanborn et al. (2013) could be easily modified to explain

Downloaded from jov.arvojournals.org on 02/02/2026

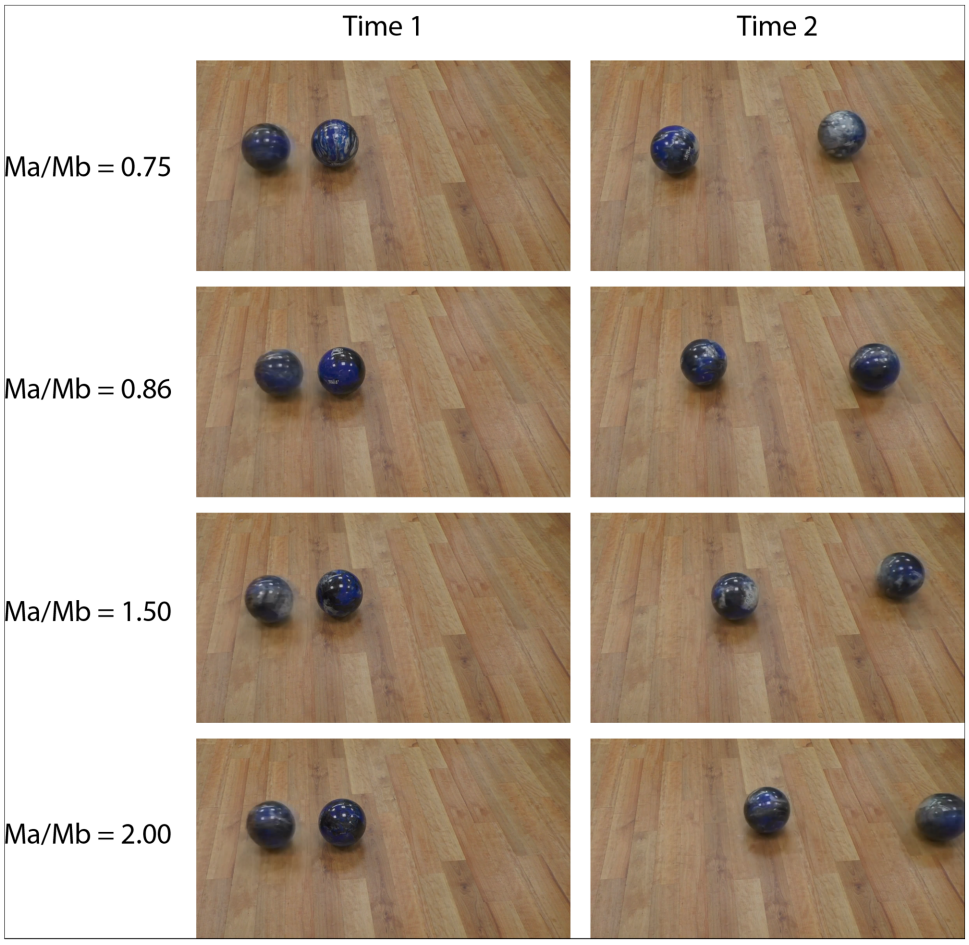


Figure 5. The instantaneous positions before and after an initially moving 12 lb. bowling ball collides with a stationary bowling ball. From top to bottom the weights of the stationary balls are 16 lb., 14 lb., 8 lb., and 6 lb., respectively. Videos of these events can be observed in Animation 4 of the supplementary materials. These materials were provided by Jason Fischer, with permission.

these results. They could of course add additional priors for friction and spin, but that would not be sufficient. They would also have to modify Equation 1 to model how the object motions are influenced by those parameters. These effects can be quite complex as shown in Animation 4. Note in particular how the moving object can slide backward after a collision while spinning in the opposite direction, and that it then reverses its direction of motion as the spin gains traction.

## Computation of relative mass for collisions among multiple objects

Another possible paradigm for studying the perceived relative mass of colliding objects has been reported by Hamrick, Battaglia, Griffiths, and Tenenbaum (2016). They presented observers with computer animations of collapsing Jenga towers that included the effects of friction and spin (see also Battaglia, Hamrick, & Tenenbaum, 2013; Mitko & Fischer, 2020). The towers were composed of rectangular blocks with two different colors that had a tenfold difference in their relative mass. Observers watched the collapse of each tower until all of the blocks had come to rest, and they were asked to judge which colored objects appeared heavier. The results revealed that observers can perform this task with 83% accuracy. Based on that finding, the authors argued that the human mind can accurately predict the dynamic interactions among objects in a collapsing Jenga tower using an internal "geometry engine" that has been endowed with a working knowledge of Newtonian dynamics (see also Bates, Yildirim, Tenenbaum, & Battaglia, 2019; Battaglia et al., 2013; Sanborn et al., 2013; Schwettmann, Tenenbaum, & Kanwisher, 2019; Ullman, Spelke, Battaglia, & Tenenbaum, 2017). However, it is not at all clear how this explains the perception of relative mass in these displays. Because the masses of each object must be specified as initial conditions to perform a simulation,

Downloaded from jov.arvojournals.org on 02/02/2026

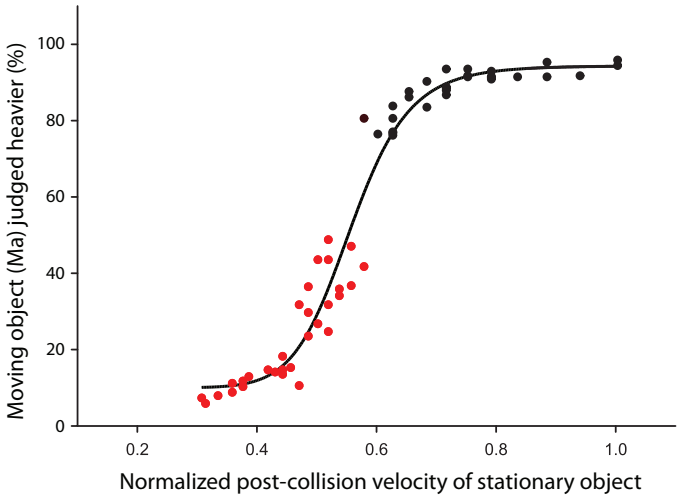


Figure 6. Relative mass judgments obtained by Mitko and Fischer (2023) for collisions between moving and stationary bowling balls. The percentage of trials that the moving object was judged as heavier is plotted as a function of the post-collision optical velocity of the stationary object. The velocities were normalized relative to the fastest one in the entire set of stimuli, so they are expressed as dimensionless numbers between 0 and 1.

that would seem to imply that the observer somehow knows the masses before the tower begins to collapse. Hamrick et al. (2016) say nothing at all about how it is possible to resolve that conundrum.

The "geometry engine" proposed by Hamrick et al. (2016) is a particularly extreme version of an internal physics model because their analysis makes no mention of any sensory inputs that are used to update its estimates before, during, or after an event unfolds. This would be especially difficult because all its essential parameters involve 3D properties of objects, but the available sensory inputs are restricted to two-dimensional optical projections.

It should also be pointed out that the 83% performance obtained in that study is highly misleading when considered in isolation. To evaluate observer sensitivity to any stimulus dimension it is necessary to consider the size of the difference they are asked to discriminate. The classic metric for measuring perceptual sensitivity was invented by Ernst Weber (1834/1996) almost 200 years ago and is now referred to as a Weber fraction. It is the ratio ($\Delta I/I$) between the intensity (I) of a standard stimulus and a just noticeable change in that intensity ($\Delta I$). If observers are 83% accurate to a tenfold difference in mass between the objects in a collapsing Jenga tower, that indicates that the Weber fraction defined by the 75% threshold would be slightly less than 10. It is important to note that is more than 100 times larger than the Weber fractions obtained for lifted weights (e.g., Norman, Norman, Swindle, Jennings, & Bartholomew, 2009; Weber, 1834/1996) or any other common stimulus dimension such as the magnitudes of light, sound, taste, or smell. It is also more than ten times larger than the Weber fraction for mass discrimination of two colliding objects that are sliding or rolling along the ground (Mitko & Fischer, 2023; Sanborn et al., 2013; Todd & Warren, 1982).

When compared to previously reported discrimination performance over a wide range of stimulus dimensions, the results obtained by Hamrick et al. (2016) show clearly that human observers are remarkably insensitive to the relative mass of objects in a collapsing Jenga tower. This makes it reasonable to question whether observers perform these judgments using an internal model of Newtonian dynamics. After all, what good is a "geometry engine" if it can barely tell the difference between a bottle of wine that weighs 3 lbs and a cinder block that weighs 30 lbs?

It is important to keep in mind that Hamrick et al. (2016) did not consider any alternative hypotheses to explain their data. Given the poor performance on their mass discrimination task, it is reasonable to suspect that observers' judgments may have been based on some form of visual information that is only roughly correlated with relative mass. For example, the lighter objects in a collapsing Jenga tower will generally have a higher average velocity, be displaced farther, and take longer to come to rest, because it takes less energy to move them. Could any of these properties provide the relevant information for judgments of relative mass in a collapsing Jenga tower? The experiment described below was designed to consider that issue.

# Methods

## Apparatus

The stimuli were displayed on an Apple iMac computer with a 24-inch 4.5K retina display (4480 × 2520 pixel resolution). The monitor was located at a 60 cm viewing distance. The luminous intensity of the monitor, measured over an area of 25°, had a minimum intensity (for black) of 1 cd/m$^2$ and a maximum intensity (for white) of 136 cd/m$^2$.

## Observers

The displays were judged by one of the authors (JFN) and 10 other observers who were completely naïve about the purpose of the experiment or how the displays were generated. All observers possessed normal or corrected-to-normal visual acuity.

Downloaded from jov.arvojournals.org on 02/02/2026

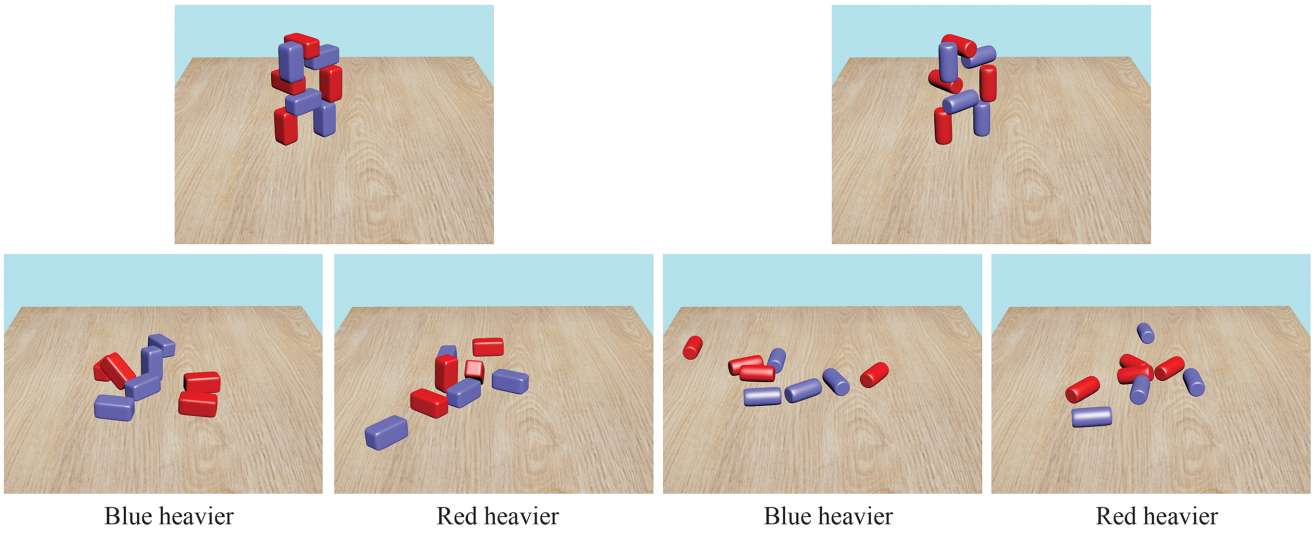


Figure 7. The first and last frames of all four versions for one of the tower configurations used in the present experiment. Videos of these events can be observed in Animation 5 of the supplementary materials.

## Stimuli

The animations were created using the Massfx physics simulator by Autodesk. The displays depicted towers composed of eight elongated objects. The depicted objects were cuboids in half of the displays, and cylinders in the remaining half, which allowed them to roll more easily. The cylinders had a length of 80 cm and a diameter of 40 cm, and the dimensions of the cuboids were 80 × 40 × 40 cm, The depicted configurations were inherently unstable and would collapse under their own weight when subjected to gravity. Half the objects were colored red, and the remaining ones were colored blue, and each of the colors was associated with a different mass that differed by a factor of ten. The coefficient of static and dynamic friction for all the depicted surfaces was 0.5, and the coefficient of restitution was also 0.5. The entire set of stimuli included 24 animations depicting six different tower configurations. Each configuration had four distinct versions with all possible combinations of cuboids versus cylinders, and red objects heavier versus blue objects heavier. Figure 7 shows the first and last frames of all four versions for one of the tower configurations, and videos of these events can be observed in Animation 5 of the supplementary materials.

Because it is not mathematically feasible to obtain closed form solutions of such complex behavior, the simulations are generated using a sequential process that calculates the dynamic interactions among objects over short time intervals. Each animation consisted of 100 individual frames with a spatial resolution of 800 × 600 pixels. The individual images were computed using the V-Ray renderer by Chaos with a camera angle of 30°, and they were presented at a rate of 30 frames/sec. The rendered images were globally tone mapped for the Apple monitor into the sRGB 2.1 color space with a D65 white-point and a gamma of 2.2. No other global histogram adjustments (e.g., tint or burn) or local sharpening or contrast enhancement operators were used.

## Procedure

On each trial, observers were presented with a computer animation for a duration of 3.33 sec followed by a blank screen. Observers were required to indicate whether the red objects or the blue objects appeared heavier by pressing an appropriate response key. An experimental session was composed of six blocks, and each block included all 24 possible animations in a random order.

# Results

Over all 11 observers, we obtained 1584 judgments of relative mass. It is important to keep in mind that the heavier objects in each display were ten times heavier than the lighter objects, yet the overall accuracy of the observers' judgments was only 57%. That finding indicates that the Weber fraction for this task was slightly above 10, whereas it was slightly below 10 for the results obtained by Hamrick et al. (2016). There were several differences between the two experiments that deserve to be highlighted. Our towers had fewer objects (eight vs. 10) and those objects had a smaller

Downloaded from jov.arvojournals.org on 02/02/2026

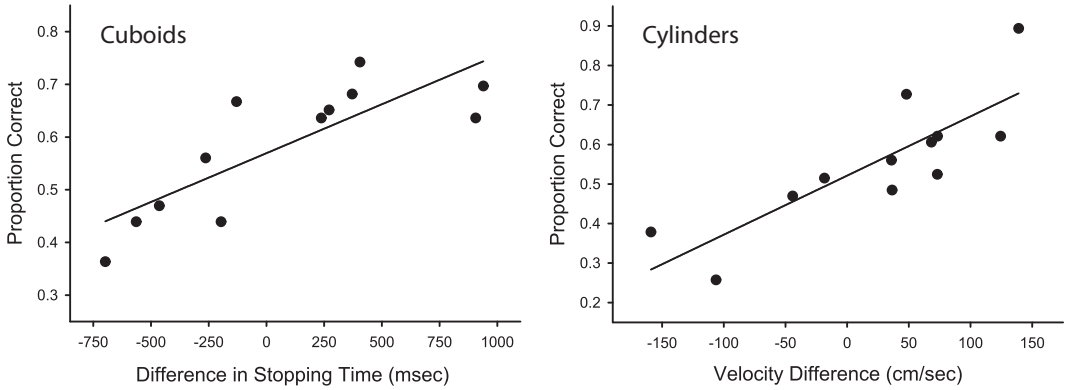


Figure 8. Left: The proportion of correct responses for the individual cuboid configurations plotted as a function of the stopping time of the lighter objects minus the stopping time of the heavier objects. Right: The proportion of correct responses for the individual cylinder configurations plotted as a function of the fastest velocity among the lighter objects minus the fastest velocity among the heavier objects at the termination of each display.

aspect ratio (2:1 vs. 3:1). The most important difference, however, is that Hamrick's observers received training with feedback, which would have allowed them to learn artifactual stimulus properties that covaried with relative mass. Our observers in contrast received no feedback at all. Nevertheless, what both experiments have in common is that the Weber fraction for discriminating relative mass in a collapsing Jenga tower is more than 100 times larger than the Weber fractions for discriminating lifted weights (e.g., Norman et al., 2009; Weber, 1834/1996) and more than ten times larger than the Weber fraction for discriminating the relative mass of two colliding objects that are sliding or rolling along the ground (Mitko & Fischer, 2023; Sanborn et al., 2013; Todd & Warren, 1982). This suggests that observers' judgments may have been based on some form of visual information that is only roughly correlated with relative mass.

In an effort to identify that information, we also analyzed the relative performance for the 24 individual displays, each of which had a total of 66 responses. The accuracy ranged from 25% to 89%. Note that this could not have occurred if the responses to individual stimuli were samples of a single underlying distribution. What this suggests is that observers' responses may have been based on sources of information that varied systematically among the different stimuli.

One possible source of information is the relative dispersion of the red and blue objects. In general, lighter objects will be dispersed farther from one another than heavier objects because it takes less energy to move them. This is clearly evident for the cylindrical objects depicted in Figure 7, but it is much less obvious for the cuboids. A regression analysis revealed that the accuracy for the individual stimuli was moderately correlated with the difference in dispersion between the red and blue cylinders ($r = 0.60$), but that the correlation was negligible for the cuboids ($r = -0.17$).

Another possible source of information is the amount of time it takes for the objects to come to a complete stop. In general, the lighter objects will stop moving later than the heavier ones. A regression analysis revealed that the accuracy for individual stimuli is strongly correlated with the difference in stopping time between the red and blue cuboids ($r = 0.81$). This analysis could not be performed on the cylinder configurations because many of the objects were still moving when the displays terminated. Thus for those stimuli we measured the difference in image velocity between the fastest red object and the fastest blue object at the end of each display, and that was highly correlated with the accuracy for the individual cylinder configurations ($r = 0.82$). Figure 8 shows scatter plots of how perceptual performance varied systematically with both of those sources of information.

Similar variations among individual stimuli were also reported by Hamrick et al. (2016). They found that the judgments for 23% of their stimuli were not significantly above chance, and they suggested that the information in those stimuli may have been inadequate for performing the required task. However, they did not elaborate about what that information might be, nor did they explain how it could have influenced the accuracy of the internal simulations. The present research has shown that the relative velocities (or durations of motion) can account for the differences among individual stimuli, but that cannot be what Hamrick et al. (2016) had in mind. Their model makes no mention of how objects interact with one another after a tower begins to collapse. Its judgments about relative mass are based entirely on the priors at the beginning of each trial.

If that were true for human observers, then it ought to be possible to judge relative mass from the initial tower configuration without seeing the tower collapse. Indeed, Hamrick et al. (2016) reported that judgments

Downloaded from jov.arvojournals.org on 02/02/2026

of relative mass are highly correlated with judgments of whether a given tower configuration will collapse. Note that this is only possible because of the rigid constraints they imposed on the construction of their towers. In general, any given tower configuration can occur with any possible combination of relative mass among its constituent objects. Without seeing the objects interact with one another, there is no way in principle to determine relative mass from an internal simulation.

## Discussion

There are two general models that have been proposed in the literature to explain the abilities of human observers to discriminate the relative mass of colliding objects. The internal physics approach involves an explicit computation of relative mass based on Newton's laws of motion (Runeson, 1977; Sanborn et al., 2013; Hamrick et al., 2016). Current implementations of that approach are limited by unrealistic assumptions about the absence of friction or spin, and they often require omniscient knowledge about the distributions of physical properties in the environment like mass, elasticity, or velocity.

An alternative sensory based explanation for the perception of relative mass is that observers' judgments are based on measurable optical variables that are partially correlated with relative mass in the natural environment (Todd & Warren, 1982; Gilden & Proffitt, 1989). We have reanalyzed the data from three previous experiments by Mitko and Fischer (2023), Sanborn et al. (2013), and Todd and Warren (1982) to explicitly evaluate that hypothesis. The results show clearly that observers' judgments are highly correlated with the post collision optical velocities of objects. We have also replicated an additional study by Hamrick et al. (2016) involving more complex collisions among multiple objects, and the results confirm that observers' judgments are most likely based on the relative optical velocities of objects or their durations of movement.

The hypothesis that mass judgments are based on relative optical velocity provides better fits to the empirical data than an explicit computation of relative mass based on Newtonian dynamics, with a much smaller number of free parameters. It is important to keep in mind that post-collision velocities (or durations of motion) are generally correlated with relative mass, but they are also affected by other factors such as elasticity, friction or spin, which can cause systematic errors in observers' judgments.

This hypothesis is also supported by the recent findings of Deeb and Domini (2025) on the effects of size on the perception of relative mass. Because different objects can vary in density, size per se should not be a relevant factor for the computation of relative mass, but the empirical results show clearly that it has a large effect on human performance. Note that this pattern of performance is quite reasonable from an ecological perspective because size is generally correlated with mass in the natural environment, even though that correlation is far from perfect.

Why would observers base their judgments of mechanical properties like relative mass on optical information that is not perfectly correlated with those properties in the natural environment? Perhaps that is the best information that is available. As we have argued repeatedly throughout this article, the internal physics models proposed for judging relative mass based on classical Newtonian dynamics all involve highly restrictive assumptions that are seldom satisfied in the natural environment, and they often use extensive prior knowledge that would be difficult to obtain through natural experience.

### Prediction of future outcomes for dynamic events

One of the main selling points of an internal "geometry engine" is that it is able to predict how dynamic events will evolve over time in the immediate future. Collapsing Jenga towers were first introduced as perceptual stimuli by Battaglia et al. (2013) to demonstrate that capability. Because it is not mathematically feasible to obtain closed form solutions of such complex behavior, the simulation of these events involves a sequential process that calculates the dynamic interactions among objects over short time intervals. Battaglia et al. (2013) argue that the perception of dynamic events involves a mental model of Newtonian dynamics that computes a similar sequential simulation to predict the unfolding patterns of motion from optical stimulation. The evidence they used to support that claim is that observers can judge the direction in which a tower will fall with a modest amount of accuracy.

One likely source of information for performing that task is the asymmetry of the initial tower configuration, but Battaglia et al. (2013) did not consider that possibility. They simply assumed that performance must have been based on an internal "geometry engine" (see also Bates et al., 2019; Hamrick et al., 2016; Sanborn et al., 2013; Schwettmann et al., 2019; Ullman et al., 2017). However, there is a fundamental problem with modeling predictive performance as if it were a computer graphics algorithm. In the context of computer graphics, all the initial conditions for the simulation are specified in advance, including the mass, density, elasticity, and friction of each object. The three-dimensional (3D) shapes and

Downloaded from jov.arvojournals.org on 02/02/2026

positions of the objects (including the occluded parts) must also be known at each time step while performing the computations. Proponents of an internal physics simulation say nothing at all about how that information could potentially be acquired from perspectively distorted two-dimensional images.

There are many examples of predictive performance in visually guided action. Consider the ability of an outfielder in baseball to determine whether to run forward or backward to catch a fly ball. One way to make that decision using an internal "geometry engine" might be to calculate the trajectory of the ball based on Newtonian dynamics combined with prior knowledge about the most likely distributions of pitch velocity, the mass and swing speed of the bat, the contact point with the ball, air resistance, and wind speed. The problem with that approach is to explain how that knowledge could conceivably be obtained.

An alternative possibility might be to calculate the trajectory of the ball by extrapolating its current movement at any given moment in time. For example, Belousov, Neumann, Rothkopf, and Peters (2016) have proposed a model that estimates the trajectory of the ball by measuring its 3D position and velocity. There are two fundamental problems with that: First, human observers do not have the sensory capacity to judge the 3D position and velocity of a falling projectile with any degree of accuracy. Second, any prediction of a ball's behavior that assumes it is traveling in a vacuum is doomed to failure. There is a large literature on the physics of baseball, and that research has shown that the effects of spin, air resistance, and wind speed can reduce the distance a ball travels by over 30% from how it would behave in a vacuum (https://baseball.physics.illinois.edu/carry.html).

Still another possibility is that catching performance is based on visual information that is highly correlated with the relative movements of the ball and the outfielder. For example, if the optical motion of the ball is accelerating upward in the visual field, that provides potential information that it will land behind the point of observation. Conversely, a decelerating optical motion indicates that it will land in front of the observer (e.g., see Chapman, 1968; Todd, 1981). This is the simplest strategy for ball catching that has been proposed to date, and it is supported by a considerable amount of psychophysical evidence (see Fink, Foo, & Warren, 2009, for a comprehensive review).

It makes no scientific sense to invoke complex computational mechanisms that depend on implausible prior knowledge as an acceptable explanation of predictive performance without making any effort to consider alternative hypotheses or to offer any explanation of how the necessary prior knowledge could potentially arise through learning or evolution. This is an area of research where Occam's razor is particularly relevant. Other things being equal, a simple explanation that relies on easily measurable optical information should be preferred over more complex explanations that require implausible prior knowledge or unlikely perceptual capabilities.

## Soft body dynamics

Thus far, our discussion about dynamic events has focused on rigid bodies whose shapes remain invariant when they interact with other objects. However, there is a large class of events that are frequently encountered in the natural environment, for which the shapes of objects can be systematically deformed. This includes the flow of liquids, the dispersion of smoke, the fluttering of a flag, or the bouncing of a water balloon. These events can all be simulated in modern computer graphics software by modeling objects or materials as large systems of interacting particles. By controlling the forces that attract and repel these particles at a microscopic level, it is possible to simulate

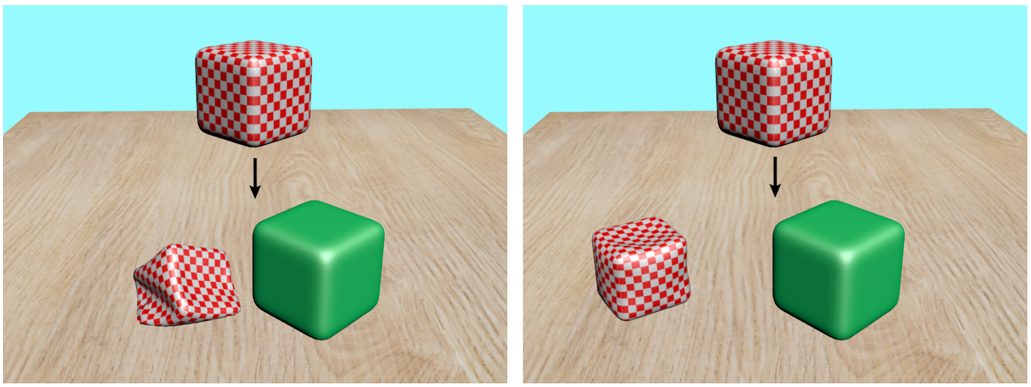

Figure 9. Schematic collisions of a falling checkered box that collides with a stationary green box. Videos of those events are shown in Animation 6 of the supplementary materials. Note that the checkered object on the right appears more inflated than the one on the left.

Downloaded from jov.arvojournals.org on 02/02/2026

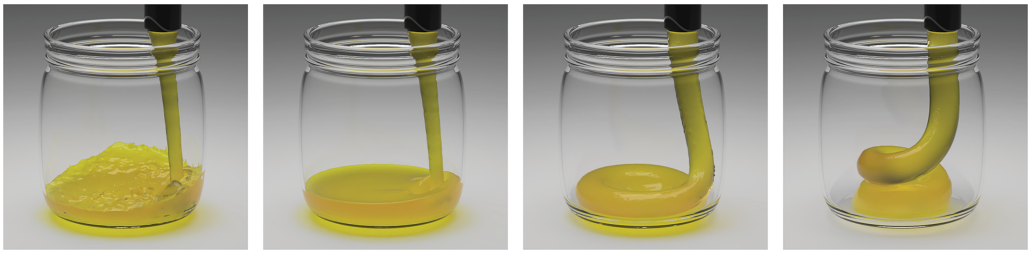

Figure 10. Still images of fluid materials with varying viscosity flowing into a glass jar from Todd and Norman (2025). Videos of those events can be observed in Animation 7 of the supplementary materials. Observers can reliably categorize these materials as juice, oil, honey, or caulk, based on the distinct shapes of the fluid surfaces.

the macroscopic behavior of a wide range of materials in a wide variety of natural contexts.

Bates et al. (2019) have proposed an internal physics model that performs a similar simulation of particle systems to explain the perception of dispersive or elastic events, but there are several fundamental problems with that proposal. How could observers possibly know that the behavior of viscous or elastic materials can be decomposed into the rigid interactions of thousands of small particles that are not visible to the naked eye, and how could they determine the many free parameters that would be necessary to set the initial conditions for such a simulation? Bates et al. (2019) say nothing at all about how it might be possible to address those issues.

A better theoretical approach, in our opinion, is to look for visually identifiable aspects of macroscopic behavior that are statistically correlated with the dynamic properties of objects or materials. Consider, for example, the two schematic collisions depicted in Figure 9, and the videos of those events in Animation 6 of the supplementary materials. Note that the checkered object on the right appears more inflated than the one on the left. Is a mental simulation really necessary to perform that judgment? A more likely explanation is that the appearance of relative inflation is based on the manner and magnitude of the 3D shape change. There have been several recent studies that have examined how the perception of softness is influenced by how objects respond to external forces, and they have all shown that shape changes are a critical source of information (Paulun, Schmidt, van Assen, & Fleming, 2017; Schmidt, Paulun, van Assen, & Fleming, 2017; Ujitoko, & Kawabe, 2022).

Similar results have been obtained for the perception of fluid dynamics (Kawabe, Maruya, Fleming, & Nishida, 2015; Paulun, Kawabe, Nishida, & Fleming, 2015; Schmid & Doerschner, 2018; Todd & Norman, 2025; van Assen & Fleming, 2016; van Assen, Barla, & Fleming, 2018). Figure 10 shows still images of fluids flowing into a glass jar, and the videos of those events can be observed in Animation 7 of the supplementary materials. Todd and Norman (2025) asked observers to classify these events as juice, oil, honey or caulk, and observers were able to do so with a high degree of reliability. The relevant visual information in this case involves diagnostic shape features, like the presence of turbulence, that only occur over limited ranges of viscosity. Although observers cannot perceive viscosity per se as it is formally defined by physicists, they can easily identify the visual effects of viscosity in their day-to-day interactions with fluid materials.

## Conclusions

The research described in the present article was designed to investigate the visual information by which human observers are able to discriminate the relative mass of colliding objects. We began by reanalyzing the data from three previous experiments by Mitko and Fischer (2023), Sanborn et al. (2013), and Todd and Warren (1982), and the results revealed that observers' judgments are most likely based on the post-collision velocities of the objects. We then replicated an additional study by Hamrick et al. (2016), which examined the perception of relative mass in collapsing Jenga towers. The results revealed that the overall accuracy of these judgments was only slightly above chance, but that the responses to individual stimuli were highly correlated with measurable optical properties like relative velocity or the relative duration of motion, which are often correlated with relative mass in the natural environment. These results provide strong evidence that judgments of relative mass can be achieved without having to simulate complex events using an internal geometry engine based on the laws of Newtonian dynamics. What they suggest instead is that dynamic event perception is primarily based on visually identifiable aspects of macroscopic behavior that are statistically correlated with the physical properties of objects or materials.

*Keywords: event perception, motion perception, 3D shape*

Downloaded from jov.arvojournals.org on 02/02/2026

## Acknowledgments

Supported by a grant from the National Science Foundation (BCS-2238179).

Commercial relationships: none.
Corresponding author: James T. Todd.
Email: todd.44@osu.edu.
Address: Department of Psychology, The Ohio State University, Columbus, OH 43210, USA.

## References

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. (2019). Modeling human intuitions about liquid flow with particle-based simulation. *PLoS Computational Biology, 15*(7), e1007210, https://doi.org/10.1371/journal.pcbi.1007210.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, 110*, 18327–18332, https://doi.org/10.1073/pnas.1306572110.

Belousov, B., Neumann, G., Rothkopf, C. A., & Peters, J. R. (2016). Catching heuristics are optimal control policies. *Advances in Neural Information Processing Systems 29*, 1–9.

Chapman, S. (1968). Catching a baseball. *American Journal of Physics, 36*, 868–870, https://doi.org/10.1119/1.1974297.

Cohen, A. L. (2006). Contributions of invariants, heuristics, and exemplars to the visual perception of relative mass. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 574–598, https://doi.org/10.1037/0096-1523.32.3.574.

Deeb, A-R., & Domini, F. (2025). The online processing of dynamics. *Cognition, 270*, 106416, https://doi.org/10.1016/j.cognition.2025.106416.

Fink, P.W., Foo, P.S., & Warren, W.H. (2009). Catching fly balls in virtual reality: a critical test of the outfielder problem. *Journal of Vision, 9*(13):14, 1–8, https://doi.org/10.1167/9.13.14.

Gilden, D. L., & Proffitt, D. R. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 372–383, https://doi.org/10.1037/0096-1523.15.2.372.

Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition, 157*, 61–76, https://doi.org/10.1016/2016.08.012.

Jacobs, D. M., Michaels, C. F., & Runeson, S. (2000). Learning to perceive the relative mass of colliding balls: The effects of ratio scaling and feedback. *Perception & Psychophysics, 62*, 1332–1340, https://doi.org/10.3758/BF03212135.

Jacobs, D. M., Runeson, S., & Michaels, C. F. (2001). Learning to visually perceive the relative mass of colliding balls in globally and locally constrained task ecologies. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1019–1038, https://doi.org/10.1037/0096-1523.27.5.1019.

Kawabe, T., Maruya, K., Fleming, R. W., & Nishida, S. (2015). Seeing liquids from visual motion. *Vision Research, 109*, 125–138, https://doi.org/10.1016/j.visres.2014.07.003.

Michotte, A. (1946/1963). The perception of causality. T. Miles & E. Miles. *Trans*. Methuen.

Mitko, A., & Fischer, J. (2020). When it all falls down: The relationship between intuitive physics and spatial cognition. *Cognitive Research: Principles and Implications, 5*, 24, https://doi.org/10.1186/s41235-020-00224-7.

Mitko, A., & Fischer, J. (2023). Do striking biases in mass inference reflect a flawed mental model of physics? *Journal of Experimental Psychology: General, 152*, 2636–2650, https://doi.org/10.1037/xge0001399.

Norman, J. F., Norman, H. F., Swindle, J.M., Jennings, L.R., & Bartholomew, A.N. (2009). Aging and the discrimination of object weight. *Perception, 38*, 1347–1354, https://doi.org/10.1068/p6367.

Paulun, V. C., Kawabe, T., Nishida, S., & Fleming, R. W. (2015). Seeing liquids from static snapshots. *Vision Research, 115*, 163–174, https://doi.org/10.1016/j.visres.2015.01.023.

Paulun, V. C., Schmidt, F., van Assen, J. J. R., & Fleming, R. W. (2017). Shape, motion, and optical cues to stiffness of elastic objects. *Journal of Vision, 17*(1):20, 1–22, https://doi.org/10.1167/17.1.20.

Runeson, S. (1977). *On visual perception of dynamic events* Doctoral Dissertation at the Faculty of Social Sciences, University of Uppsala.

Runeson, S., & Andersson, I. E. (2007). Achievement of specificational information usage with true and false feedback in learning a visual relative mass discrimination task. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 163–182, https://doi.org/10.1037/0096-1523.33.1.16.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review, 120*, 411–437, https://doi.org/10.1037/a0031912.

Downloaded from jov.arvojournals.org on 02/02/2026

Schmid, A. C., & Doerschner, K. (2018). Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials. *Journal of Vision, 18*(1):14, 1–32, https://doi.org/10.1167/18.1.14.

Schmidt, F., Paulun, V. C., van Assen, J. J. R., & Fleming, R. W. (2017). Inferring the stiffness of unfamiliar objects from optical, shape, and motion cues. *Journal of Vision, 17*(3):18, 1–17, https://doi.org/10.1167/17.3.18.

Schwettmann, S., Tenenbaum, J. B., & Kanwisher, N. (2019). Invariant representations of mass in the human brain. *eLife, 8*, e46619, https://doi.org/10.7554/eLife.46619.

Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception & Performance, 7*, 795–810, https://doi.org/10.1037/0096-1523.7.4.795.

Todd, J. T., & Norman, J. F. (2025). The identification of materials from patterns of fluid flow. *Journal of Vision, 25*(10):11, 1–9, https://doi.org/10.1167/jov.25.10.11.

Todd, J. T., & Warren, W. H., Jr. (1982). Visual perception of relative mass in dynamic events. *Perception, 11,* 325–335, https://doi.org/10.1068/p110325.

Ujitoko, Y., & Kawabe, T. (2022). Perceptual judgments for the softness of materials under indentation. *Scientific Reports, 12*, 1761, https://doi.org/10.1038/s41598-022-05864-x.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences, 21*, 649–665, https://doi.org/10.1016/j.tics.2017.05.012.

van Assen, J. J. R., & Fleming, R. W. (2016). Influence of optical material properties on the perception of liquids. *Journal of Vision, 16*(15):12, 1–20, https://doi.org/10.1167/16.15.12.

van Assen, J. J. R., Barla, P., & Fleming, R. W. (2018). Visual features in the perception of liquids. *Current Biology, 28*, 452–458, https://doi.org/10.1016/j.cub.2017.12.037.

Weber, E. H. (1834/1996). E. H. Weber on the Tactile Senses translated into English by H E Ross, D J Murray. [East Sussex, UK: Erlbaum (UK); London: Taylor & Francis] [Original work published 1834].

Downloaded from jov.arvojournals.org on 02/02/2026